February 2021
Geoff Huston

Notes from the DNS Privacy Workshop at NDSS 2021

For many years the consuming topic in DNS circles was that of the names themselves. If you wind the clock back twenty years or so you would find much discussion about the nature of the Internet's name space. Why were there both generic top-level labels and two-letter country codes. If we were going to persist with these extra-territorial generic country codes in the name space, then how many should there be? Who could or should manage them? And so on. And on. At the same time there was the DNS as a name publication vehicle and an associated name resolution protocol. That part of the DNS seemed to be so simple that it did not merit much attention at all, or so it seemed.

But the DNS is a lot like chess. It's a simple game in terms of the rules, but phenomenally complex in the way it can be played.

There are many issues that lurk just below the surface of the DNS, but in recent years one topic has been very prominent, namely privacy. You see the DNS is a precursor to almost everything we do on the Internet. And this is valuable information.

The Internet itself has largely followed the same path as newspapers and broadcast radio and television: These days, the Internet is largely an advertising platform. In advertising, as any advertiser will tell you, knowledge of the consumer is paramount. The nirvana for an advertiser is to deliver their message to consumers who can be persuaded to purchase their product. The best advertisement is a helpful and timely suggestion to a consumer. So, the Internet can tell an onlooker a lot about each and every user. But the DNS can do a whole lot more. The DNS is in a unique position to know what a user is about to do, as well has what has happened in the past. And, as the Snowden files pointed out, the circle of onlookers with an interest in this form of information leak expanded considerably.

Quite predictably, this situation has provoked a response from the various folk who build and operate the Internet's infrastructure. From the perspective of the DNS, it was clear that the open and promiscuous nature of the DNS was being widely abused. The topic of DNS Privacy was coined to cover a few issues concerning protection from eavesdropping, authenticity of the data being passed in DNS transactions and assurance as to the identity at the other end of a DNS transaction.

There have been a number of efforts to improve the DNS from a privacy perspective. DNS Query Name minimization is intended to limit the extent to which query names are broadcasted by modifying the name resolution process to expose only those parts of the query name that are relevant in each step of the name resolution task. DNS over TLS (DoT) are intended to reduce the opportunity for eavesdroppers to "see" DNS queries and responses and confirm the authenticity of the other end of a DNS transaction. And DNSSEC is there to prevent the client from being misled by a tampered DNS response.

Where now?

And that is the context of the 2021 NDSS DNS Privacy Workshop. What re the current research topic in the area of DNS Privacy?

Oblivious DNS over HTTPS (ODOH)

Let's start with the original Oblivious DNS concept (ODNS). A research group has come up with an approach to break through this uncomfortable compromise of having to admit at least some external party to being privy to both my identity and the queries that I make. The approach is one that they termed "Oblivious DNS" (written up as an Internet Draft, July 2018, also a paper).

The concept is delightfully simple, and it is intended to prevent the recursive resolver from knowing both the identity of the end point stub resolver and the queries that they are making. A one-off session key is generated by the stub resolver. This stub then takes the original query name and encrypts it using the session key. The session key is encrypted using the public key of the target ODNS server, and appended to the encrypted query name. This string is then encoded into an ascii string and treated like a domain name label itself. The querier then appends the label of an oblivious DNS server domain.

> In the DNS the QNAME field consists of 4 sets of 63 bytes, which limits both the key size and encryption scheme used. For this reason, ODNS uses 16-byte AES session keys and encrypts the session keys using the Elliptic Curve Integrated Encryption Scheme (ECIES). Once the session key is encrypted, the resulting value takes up 44 bytes of the QNAME field.

The stub passes this query as conventional DNS query to its "normal" recursive resolver. The recursive resolver is unaware of oblivious DNS label encoding and treats the query as it would any other. The recursive resolver will then pass a query for this name to an ODNS authoritative server as specified in the query name. To resolve this name the ODNS authoritative server will decrypt the session key (as it has the matching private key to decrypt the query name) and then use this session key to decrypt the query name. It can then use a conventional recursive resolution procedure to resolve the original query name. The response is encrypted using the session key. The ODNS server will then respond to the recursive resolver with the encrypted query name in the query section and the encrypted answer section that it has just generated. Upon receipt of this response, the recursive resolver will pass this to the stub resolver. The stub resolver uses its session key to decrypt the response.

The ODNS server is aware of the query name, but it is unaware of the identity of the stub resolver, as it only knows the identity of the recursive resolver. The recursive resolver is aware of the identity of the stub resolver, but is unaware of the query name.

Oblivious DOH takes this approach and builds upon it in a couple of ways. Rather than trying to operate the entire mechanism by encoding the query name and the session key into base32 encoded form of a new query name, the specification of Oblivious DOH (https://tools.ietf.org/html/draft-pauly-dprive-oblivious-doh-03) encrypts the entire original DNS query using the public key of the target ODOH server, as well as the session key. ODOH cannot use a conventional recursive resolver as an intermediary as the query itself is now encrypted. Instead, ODOH uses a ODOH Proxy that functions in a manner similar to a conventional forwarder but passes the DOH query (which as this point is still an opaque object) to the query-specified target ODOH server.

The differences between ODOH and ODNS is that ODOH appears to dispense with session keys and uses public keys for both the client and the server. The query is encrypted using the ODOH target public key, and the response is encrypted using the client's public key, as provided in the query. The target is not appended to the query name in plaintext, but instead is specified in the :path element of the URI template of a DOH query.

A report to the workshop was on the topic of measuring the performance of ODOH using a set of 90 clients evenly spread across 9 vantage points with a query load of some 15 requests per minute.

(http://bit.ly/2OBapFv). Interestingly, the major latency time consideration was HTTPS connection reuse. It also appeared to make a lot of sense, in terms of reduced latency, to use a ODOH proxy that is roughly on the same network path between the user and the target ODOH server. The ODOH Target was configured as a stub resolver to resolve the name, and the best performance outcomes were achieved when the ODOH target is co-located with the target's chosen recursive resolver, or ODOH target support was added to a DOH recursive resolver.

The major consideration here is that the choice of a ODOH Proxy and OSOH Target is an important performance factor, but even more important is the reuse of HTTPS sessions across many queries, to avoid the considerable overheads of TCP and TLS session establishment (Figure 1)

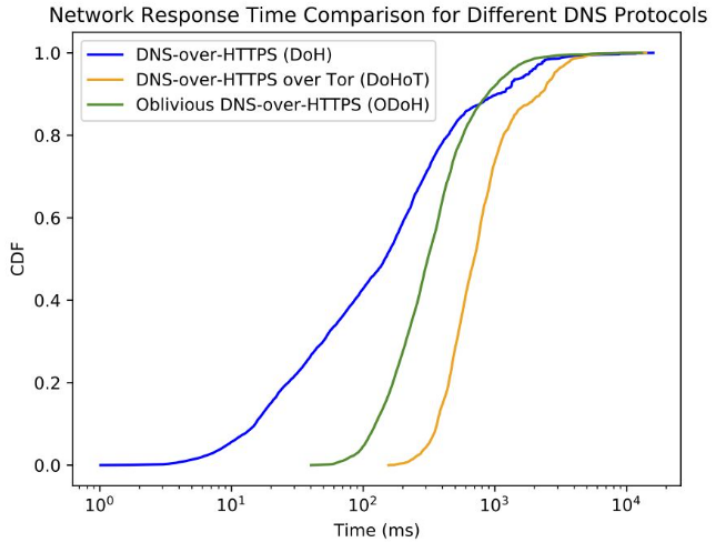

Figure 1 – Comparing ODOH with Other DNS Protocols - http://bit.ly/2OBapFv

Measuring DoT/DoH Blocking

Both DoT and DoH can hide the DNS queries from eavesdroppers, but they can't hide the fact that the user is using TLS to perform DNS queries. Simone Basso of the Open Observatory of Network Interference presented on a report that measured the level of blocking of DoT and DoH in a small set of economies (Kazakhstan, Iran and China) where such DoT and DoH blocking has been suspected to occur (http://bit.ly/3bt74kP).

DoT uses TCP port 853, and blocking of certain servers was consistent, apart from 1.1.1.1:853 in KZ, which appeared to be blocked and unblocked frequently.

DoH uses TCP port 443 and is not so readily identifiable as a DNS transaction, particularly if its use is combined with a NAT-like in-line proxy server is used on the external side. However deep packet inspection of the TLS session will reveal the server name in the SNI field of the TLS handshake, and this was observed to occur. The deployment of ECH (Encypted Client Hello) is still in the early stages of deployment, which inplies that SNI inspection is still capable of revealing details of the intended DNS server even if in-line proxies are being used.

Speed isn't Everything

It is a common wisdom in DNS circles that speed is of paramount importance. Why did we largely reject putting our WebPKI keys in the DNS and securing this with DNSSEC? Because DNSSEC validation is so tediously slow, we say to each other, and applications, and users are intolerant of sluggish systems. "Every millisecond matters," we say to each other. When we evaluate new technologies in the DNS, we evaluate them against their speed, such as with the Oblivious DoH latency study mentioned already. But is absolute speed necessarily the only thing that matters? Alec Muffet argues that this speed argument is "a presumptuous act of tech privilege" (http://bit.ly/2MXIBuj).

Figure 2 – Tech Privilege - http://bit.ly/2MXIBuj

The report describes Alec's use of DOH using a Tor substrate, which he argues provides a more embracing security cloaking through multiple layers of mutually independent proxies. The inevitable cost is a higher transaction latency for such DNS transactions, but the trade-off is a more robust security configuration that has fewer points of potential privacy vulnerability.

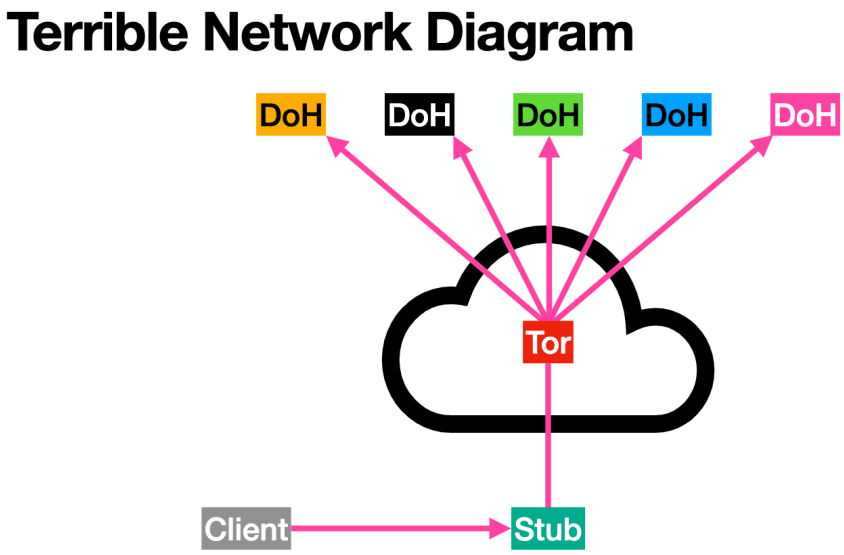

Figure 3 – DoHoT - http://bit.ly/2MXIBuj

Protocol Design Trade-offs in the Public Interest

All protocol design is inevitably an exercise in design trade-offs. Size, speed, practicality, reusability and security are all variables in this space, and the consideration of these issues is usually one of technical dimensions. These days the Internet is indeed the communications substrate for the entire world and inevitably we have to consider the public interest in these trade-offs argues Mallory Knodel of the Center for Democracy and Technology (https://bit.ly/3qnZJak). The presentation looked at various aspects of DNS privacy from a public interest perspective.

The increasing use of encryption makes measurement more challenging. In the days when the Internet was largely a research experiment, measurement played a big role in sharpening our understanding of how these systems interacted. Measurement continues to play a big role in today's Internet, not only relating to performance and efficiency but also the capability of bringing to light behaviours that might impact human rights or assumptions relating to privacy. However, the more we cloak transactions from direct sight the less we can readily observe.

As we make transactions more private, are we also allowing abuse to thrive? How can we distinguish between "normal" and "abnormal" transactions when the nature of the transactions themselves is no longer visible? Can such privacy tools be used by malicious actors to behave in ways that evade detection? If abuse can thrive in a way that evades effective attribution and even mitigation, then are we trashing the Internet and rendering it a toxic and useless wasteland?

Various forms of content censorship exist in most parts of the Internet (which itself is a claim explored in some considerable detail in another workshop presentation that I won't cover here: "Censored Planet: An Internet-wide, Longitudinal Censorship Observatory," http://bit.ly/3kYvjdI). Whether it's offensive, illegal or dangerous content, various forms of censorship, particularly through the use of DNS filters and blocks, exists in many parts of the Internet. The more DNS privacy measures evade such mechanisms the more tempting is the blanket option of complete national shutdown of Internet services in times of civil unrest. By making it more challenging to implement such measures with limited side effects are we encouraging resorting to blunter measures of shutdown?

Obviously, this was not a presentation that had all the answers, or even any of the answers, but the questions are definitely useful questions to ask in the context of making the DNS more obscure and more private.

The User Perspective

It is commonly asserted that all this work on DNS privacy is about providing protections for the user. Nick Feamster reported on a survey that was intended to test the proposition that users understood the issues and made DNS privacy settings in their applications based on such an understanding (http://bit.ly/38ijfPp). In an admittedly very small sample set, the survey asked the participants if they understood these DNS settings, why they made particular choices in these settings and the information that they used to reach these decisions and the level of trust.

The outcomes of the survey reinforce the common perception that most users do not fiddle with the settings in their devices and applications and have no real information about how to do so to enhance their personal privacy.

The conclusion appears to be that if we want to shift the dial on privacy then the default settings need to change, and at that point the underlying commentary is that "we have your best interests at heart here, even if you don't really appreciate what your bests interests might be. Trust us!" As Nick pointed out: "People don't understand the ramifications of selecting different options". It's not exactly a reassuring story if the explanation of all of this effort is in response to legitimate users' concerns over the erosion of privacy through this technology. Is this another instance of Alec Muffet's characterisation of a "presumptuous act of tech privilege?"

The Public Interest Perspective

There are a set of rather uncomfortable conversations going on in all this to understand where true public interest lies. There is little doubt that we all use the so-called "free" services where the actual revenue stream is derived from advertising revenue which itself is derived from an embracing knowledge of each and every user. On the other hand, there is little in the way of informed consent going on here, or even the option to withhold consent. The data harvesting is occurring at a deep level, such as the DNS, and the profiling of the user appears to be at a level of advertising directed to "people like me" as distinct to advertising being directing to "me", thought this tool might change in the future. The side-effects are chilling in that the industry is consolidating into a handful of extremely large entities who are now so dominant that they have the ability to not only set the terms and conditions for their competitors, such as there are any left, but also these consolidate giants appear to have the social power to define their own regulations and controls. This has never ended all that well for our society and our social institutions in the past and there is no reason to think that today's situation is any different.

However, a balance needs to be struck. If we anonymised the Internet to the extent that individual user actions did not leak any useful profile information and online advertising had about as much efficacy in terms of targeting as a daily newspaper of bygone times, would there still be a rich and valuable Internet ecosystem populated by services provided without direct cost to users? Probably not. Would the metered pay-per-use and pay-per-view economy of such an Internet be as rich, as vibrant and as useful as what we have today?

On the other hand, it is a legitimate source of concern that this industry has been way too cavalier with personal information and providing incentives to the industry to treat personal data with due respect and care seems to be the right thing to do.

But who should articulate the public interest? And how? The Internet Engineering Task Force jumped on the Snowden information and decried the practice of such insidious levels of data collection by state actors by claiming in RFC 7258 that "Pervasive monitoring is a technical attack that should be mitigated in the design of IETF protocols, where possible." However, it is difficult to pick apart the sense of emotional outrage and betrayal on the part of a cadre of the technically privileged from a considered view of what is in the best interests of a robust, balanced and resilient social framework.

Two things are clear, however. One is that is that the technical topic of DNS Privacy is one of the critical components of this discussion about information leakage and reasonable expectations of personal privacy. The second is that I suspect that the discussion is not over by any means. It has hardly begun!

Disclaimer

The above views do not necessarily represent the views or positions of the Asia Pacific Network Information Centre.

Author

Geoff Huston AM, M.Sc., is the Chief Scientist at APNIC, the Regional Internet Registry serving the Asia Pacific region.

www.potaroo.net